Title	線形予測残差を利用したスペクトル拡散型音声情報ハイディング法
Author(s)	並河，伶弥
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15905
Rights	
Description	Supervisor:鵜木　祐史，先端科学技術研究科，修士（情報科学）

Speech information hiding method based on spread spectrum
using linear prediction residue

1710151 Namikawa Reiya

The rapid development of advanced information communication technology (ICT) has positively impacted digital speech communication. However, misuse of ICT causes speech tampering and speech spoofing in digital speech communication.

Therefore, audio information hiding (AIH) techniques have recently been focused on as state-of-the-art techniques enabling secure and safety communication. To protect digital audio content, AIH techniques aim at embedding security codes as watermarks, which are inaudible and inseparable by users, and at robustly detecting embedded codes from watermarked signals against any kind of processing and various attacks. Therefore, AIH techniques generally must satisfy two important requirements: inaudibility and robustness.

The direct spread spectrum (DSS) method is a robust AIH method. A watermarked signal $y(n)$ is made by that spreads the message signal $m(n)$ by a pseudo-random noise (PN) signal $c(n)$ and adds it to the host signal $x(n)$. The PN signals have properties of $E\{c(n)\} = 0$ and $E\{c^2(n)\} = 1$, where $E\{\cdot\}$ is the expected value of "$\cdot$". Since $x(n)$, $y(n)$, and $c(n)$ are assumed to be ergodicity, $E\{\cdot\}$ can be regarded as the temporal average of "$\cdot$". Hence, $E\{y(n)c(n)\}$ used for judgment is made by multiplying the watermarked signal $y(n)$ with the same PN signal $c(n)$. This process can be used to extract message "0" or "1" from $y(n)$.

The PN signal has a white spectrum, so the watermarked signal has been spread over a wide band. Since this feature distorts the watermarked signal, the DSS method has a problem in terms of inaudibility due to the spread spectrum.

In this research proposes a speech watermarking method that can satisfy inaudibility and robustness simultaneously. A linear prediction (LP) scheme for speech analysis and synthesis is used in the proposed method. On the basis of the principle of the DSS method, the proposed method spectrally spreads a message by using LP residue and embeds the spread spectrum of the message into the host signal.

Linear predictive coding (LPC) is one of the most basic speech coding methods using linear prediction (LP). LPC generates an LP coefficient corresponding to the spectral envelope and an LP residue corresponding to the sound source of the speech signal. Similar to PN, LP residue $r(n)$ has $Er(n) = 0$ and $Er^2(n) = 1$. Therefore, in the proposed method, LP residue is used to spread the message signal.

The LP residue is used to spread the message in the host signal. The one of advantage LP residue has a similar spectral band to the host signal, so the message is spectrally spread into a similar spectrum to the host signal. Therefore, the sound distortion due to the spread spectrum will be reduced by using LP residue. In addition, the proposed method will have the high robustness, because proposed method is based on DSS method as robust AIH.

In the proposed method, parameter a can be used to set the embedding strength for watermarked signal. The parameter a is determined by the power level of the host signal L_{PHS}, the power level of the watermark signal L_{PWS}, and the strength setting level L_{SSL}. The L_{SSL} was determined by evaluations of sound quality distortion (SQD) and the bit error rate (BER). The BER and SQD tests were carried out to determine the setting value. In the BER tests, BER of 10% was set as the criterion. In the SQD tests, ; log-spectrum distortion (LSD) and perceptual evaluation of speech quality (PESQ) were used in the evaluation method. In these tests, LSD of 1 dB or less and PESQ of -1 ODG or higher were used as criteria for sound distortion. The objective difference grades (ODGs) were graded on a five-point scale as 0 (imperceptible), -1 (perceptible), -2 (slightly annoying), -3 (annoying), and -4 (very annoying). Evaluation results, The L_{SSL} was determined to be -10 dB when the payload was 4 bps. Then, the mean L_{all} was -20.6 dB. When the payload was 8 bps, there was no optimal L_{all}. When we relaxed the evaluation standard of SQD, the L_{SSL} was determined to be 0 dB, Then, the mean L_{all} was -14.0 dB.

In information detection, when the host signal and the watermarked signal were not synchronized, accurate information detection is difficult. Therefore, synchronizes the two signals was required. In frame synchronization, correlation was taken while sliding two signals, and the point where the correlation value became maximum was taken as the synchronous position. The information detection rate after frame synchronization was evaluated. In the detection evaluate, BER of 10% was set as the criterion. As a detection result after signal synchronization, the BER was 2%.

We investigated whether the proposed method could satisfy the requirement of inaudibility and robustness. The inaudibility was evaluated by using the sound quality distortion (LSD and PESQ) of the watermarked signal. Robustness was evaluated by using BERs against various attacks; downsampling (44.1 to 16 kHz), addition of white Gaussian noise (WGN, signal-to-noise ratio (SNR) of 36 dB), re-quantization (RQZ, 16 bit quantization to 8 bit quantization), and speech coding (G.711, G.726, and G.729).

In these evaluations, 12 speech signals in the ATR database (B set) were used. Only the utterance sections in the speech signals were used as the stimulus. The payloads were 4, 8, 16, 32, and 64 bps. Messages were random bit streams. No error correction schemes were used.

The LSDs were under 1 dB in the proposed method, whereas the LSDs were at or above 3 dB in the DSS method. LSD by the proposed method was improved by 2 dB in comparison with that by DSS method. The PESQs exceeded -1 ODG in the proposed method, whereas the LSDs was always about -4 ODG in the DSS method. PESQ by the proposed method was improved by 3 dB in comparison with that by DSS method.

As results of the BERs of the watermarked signals against attacks, the proposed method was robust against several attacks. The proposed method has the same degree of robustness as the DSS method.